Time Series and Dynamic Systems Analyses in Drug Discovery

PS-BiOmics&Pathology-BEDA

Roche Pharma Research and Early Development, Pharmaceutical Sciences, Roche Innovation Center Basel

Tony Kam-Thong on behalf of the Bioinformatics and Exploratory Data Analysis team

Guest Lecture Mathematical and Computational Biology in Drug Discovery
University of Basel
7th of May 2021

Outline

- Time series data analysis
  - Disease progression and response to treatment
  - Mixed effects model for repeated measurements
- Dynamic systems
  - Background knowledge
  - RNA velocity estimate from Single Cell RNASeq data
    - Going beyond discrete cell types by inferring cell states and their transitions
- Disclaimer
  - Non-exhaustive list of examples
  - Many others: Stochastic processes, time to event analysis, PKPD …

Background

Drug Discovery Today • Volume 25, Number 3 • March 2020 REVIEWS

ELSEVIER

Teaser It is high time to integrate computational models at multiple levels to understand drug mechanism and safety.

Reviews • KEYNOTE REVIEW

Multiscale modelling of drug mechanism and safety

Jitao David Zhang[1,2], Lisa Sach-Peltason[1], Christian Kramer[1], Ken Wang[1] and Martin Ebeling[1]

[1] Pharma Early Research and Development, Roche Innovation Center Basel, F. Hoffmann-La Roche, Grenzacherstrasse 124, 4070 Basel, Switzerland
[2] University of Basel, Department of Mathematics and Computer Science, Spiegelgasse 1, 4051 Basel, Switzerland

- Present along the multiscale modelling path

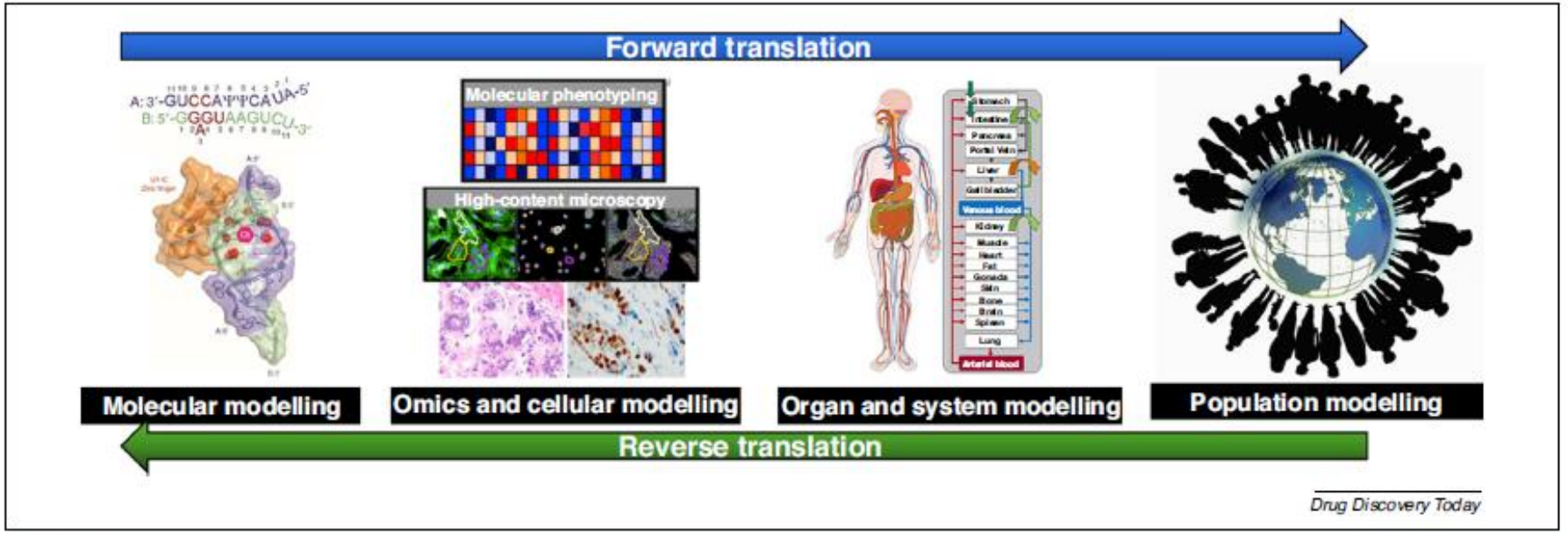


FIGURE 5

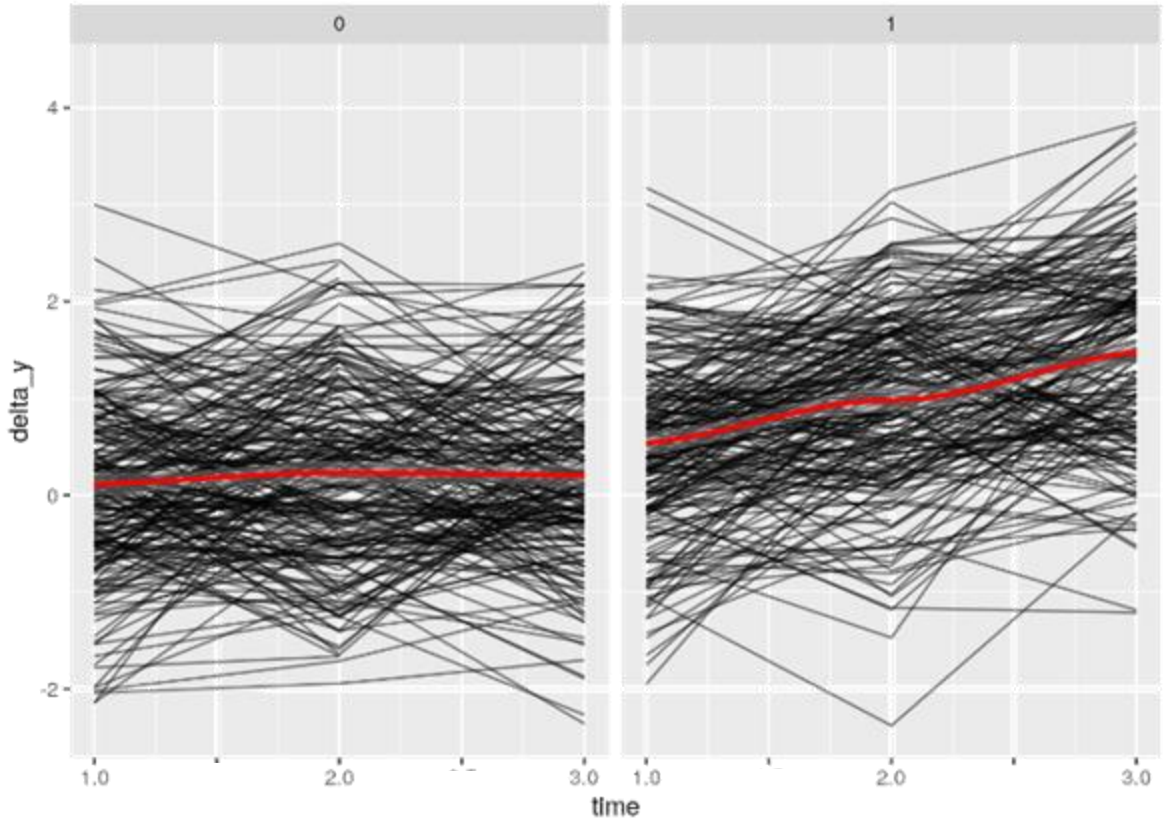
https://doi.org/10.1016/j.drudis.2019.12.009

Background – Time series data analysis

Work from Zhiwen Jiang

- Many of the data that we encounter include a temporal component and are inherently part of a dynamic system
  - Disease progression and drug response: Readout from patients over time → Mixed effects Model for Repeated Measurements (time is treated as ordinal categorical variable)

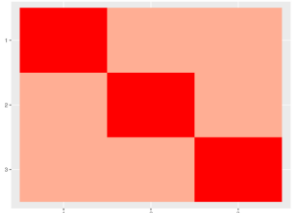
- Simulated change from baseline



- Examples of within-subject covariance structures (between 3 time points)

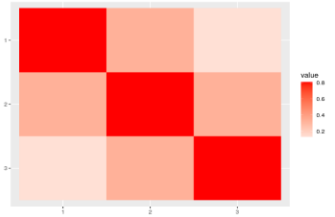
Compound Symmetry

$$\begin{bmatrix} \sigma^2+\sigma_1^2 & \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma^2+\sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 & \sigma^2+\sigma_1^2 \end{bmatrix}$$



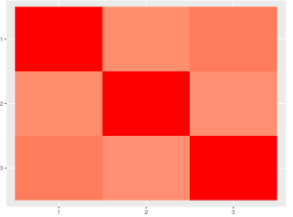
Autoregressive

$$\sigma^2\begin{bmatrix} 1 & \rho & \rho^2 \\ \rho & 1 & \rho \\ \rho^2 & \rho & 1 \end{bmatrix}$$



Unstructured

$$\begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 \end{bmatrix}$$



Many others …

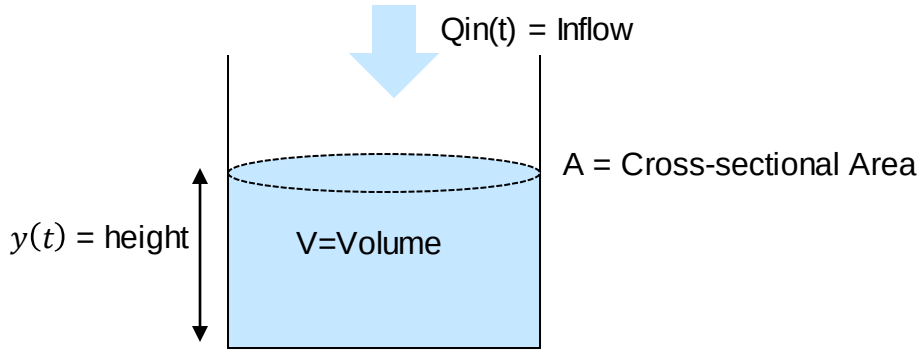
Background-MMRM Model for clinical drug response

Work from Zhiwen Jiang

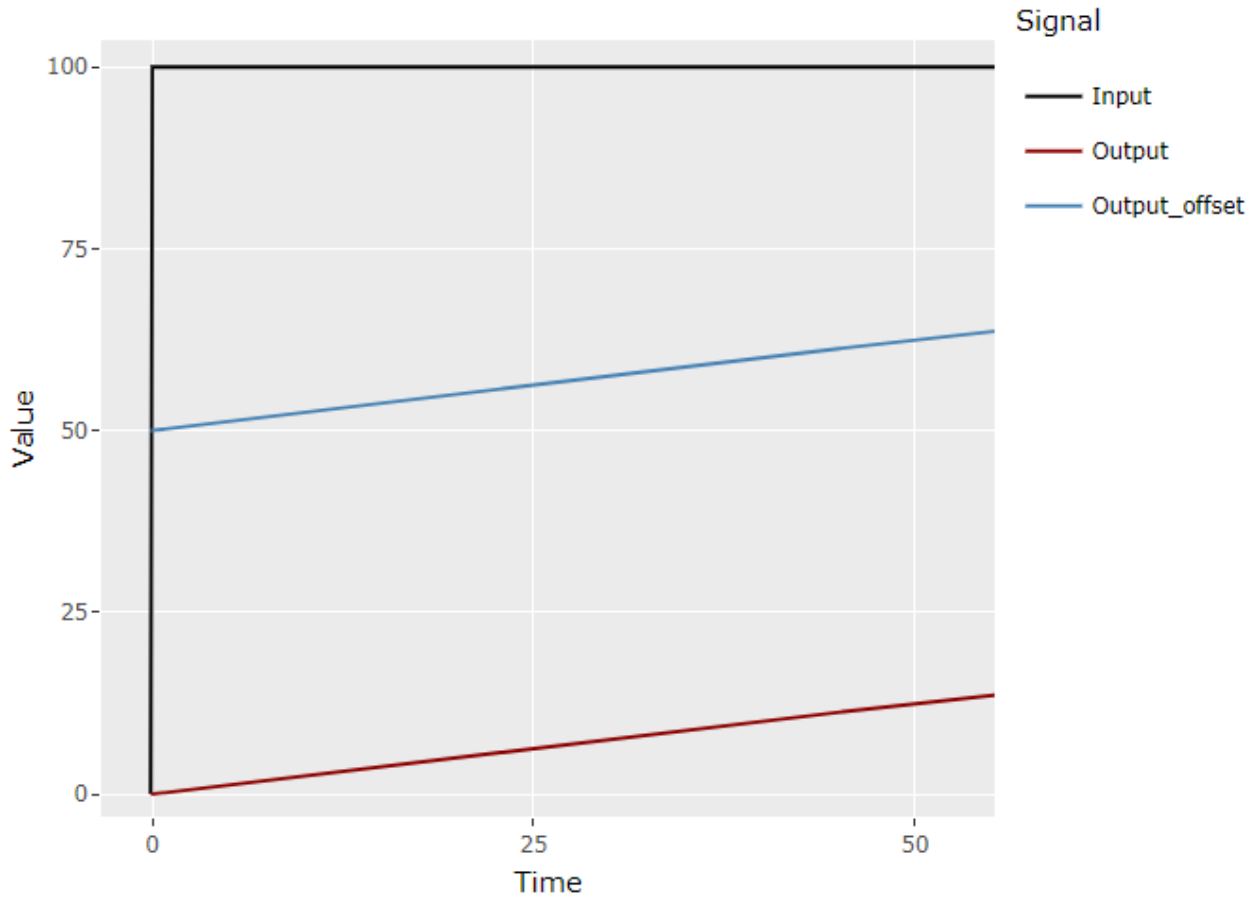
- Data can be noisy, sparse (low sampling rate) and missing
- Ignoring or misspecification of the covariance structure can lead to incorrect/inflated statistical significance of the fixed effects (time, treatment and time:treatment interaction)

Background – Dynamic Systems: Illustrative example

- Filling a water bucket

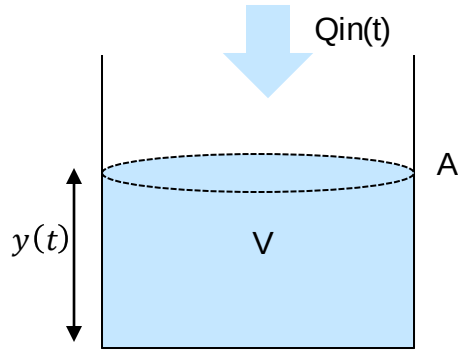


- Rate of change of volume = inflow rate
- $\frac{d(V(t))}{dt} = \frac{d(Ay(t))}{dt} = A\frac{d(y(t))}{dt} = Q_{in}(t)$
- $y(t) = y_{t0} + \frac{G}{A}\int_0^t u(\tau)\,d\tau = y_{t0} + \frac{G}{A}t$

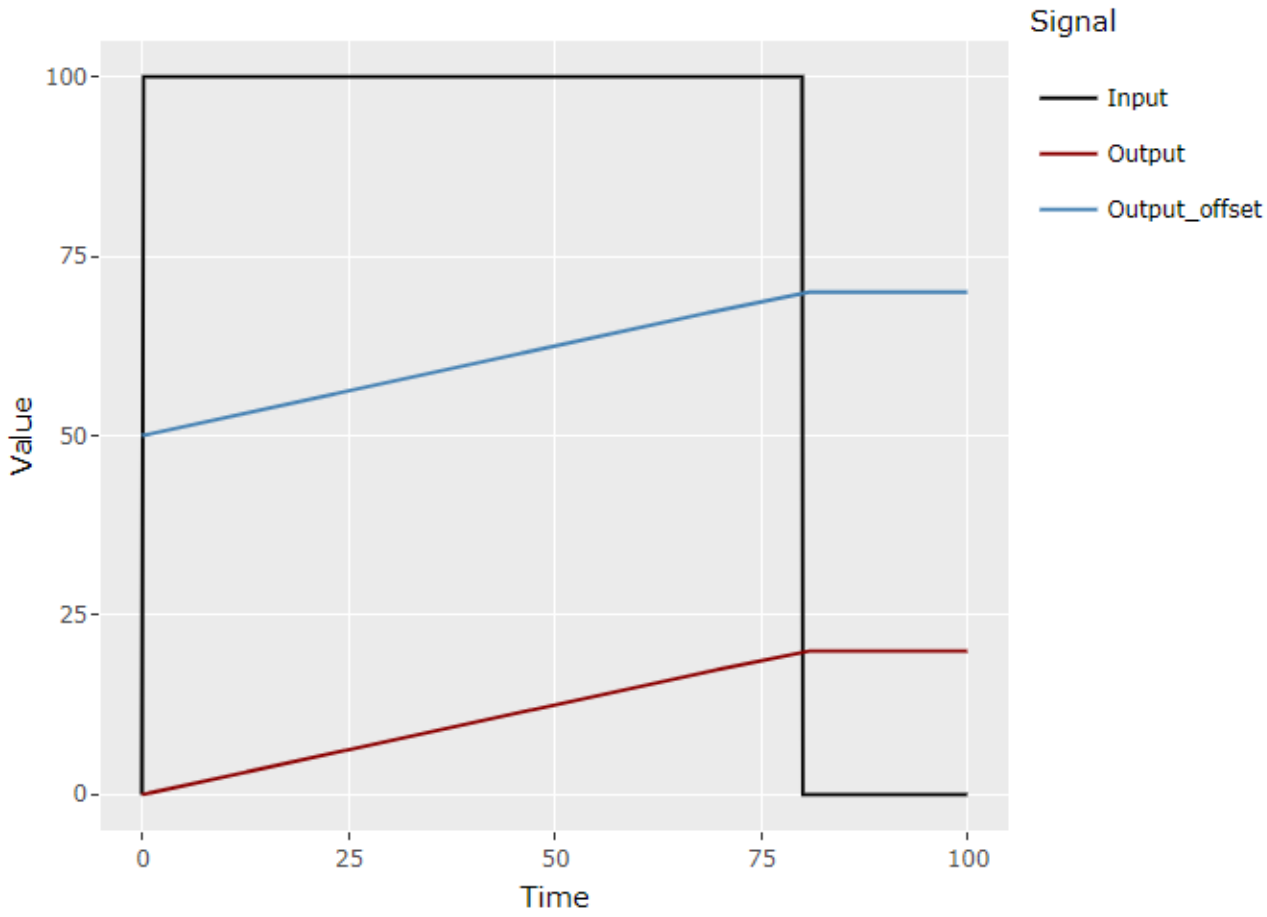


Background – Dynamic Systems

- Filling a water bucket (off switch)



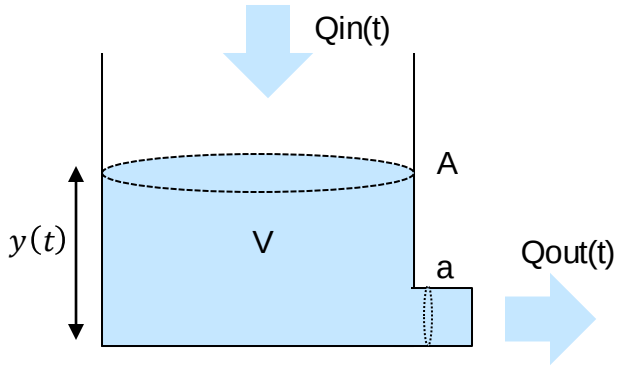
- $\frac{d(V(t))}{dt} = \frac{d(Ay(t))}{dt} = A\frac{d(y(t))}{dt} = Q_{in}(t)$
- $y(t) = y_{t0} + \frac{G}{A}\int_0^t u(\tau) - u(\tau - t_{off})\ d\tau$
- $y(t) = y_{t0} + \frac{G}{A}(\int_0^{t_{off}} u(\tau) - u(\tau - t_{off})\, d\tau + \int_{t_{off}}^t u(\tau) - u(\tau - t_{off})\, d\tau)$
- $y(t) = \begin{cases} t < t_{off}: y_{t0} + \frac{G}{A}t \\ t \geq t_{off}: y_{t0} + \frac{G}{A}t_{off} \end{cases}$



- *Constant rise (accumulation, no steady state) until tap switched off*
- *Not accounting for spills, overflow and evaporation!*

Background – Dynamic Systems

- Leaky bucket

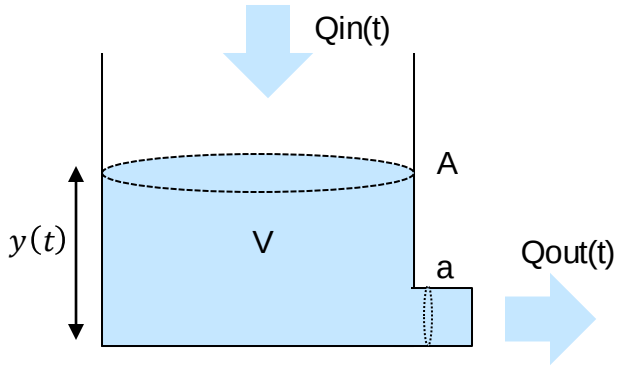


- Rate of change of volume = inflow rate – outflow rate
- $A\frac{d(y(t))}{dt} = Q_{in} - Q_{out} = Q_{in} - f(\mathrm{y(t)}) = Q_{in} - \mathrm{a}\sqrt{2g}\,y(t)$

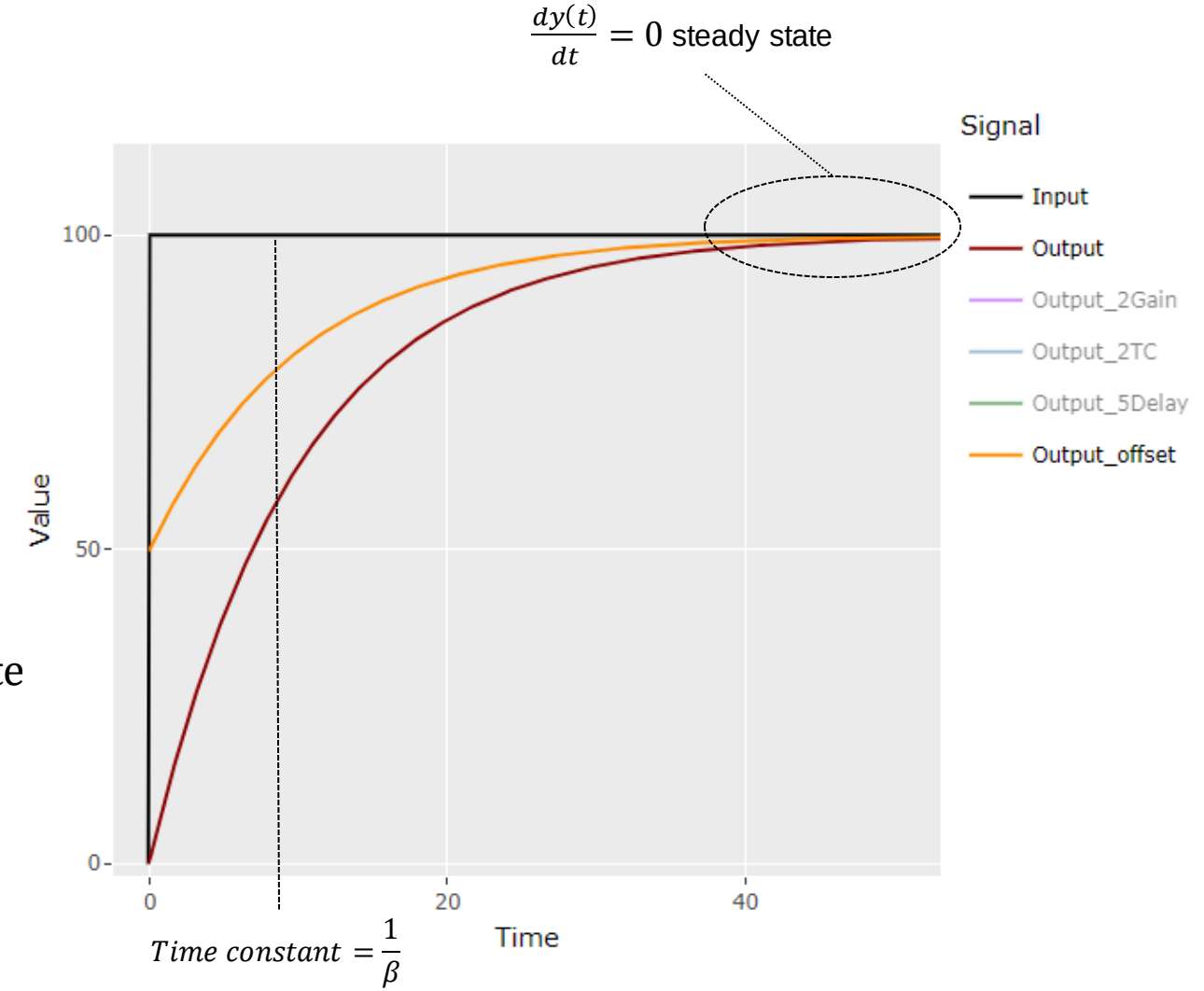
Simplification for context relevance

Background – Dynamic Systems

- Leaky bucket

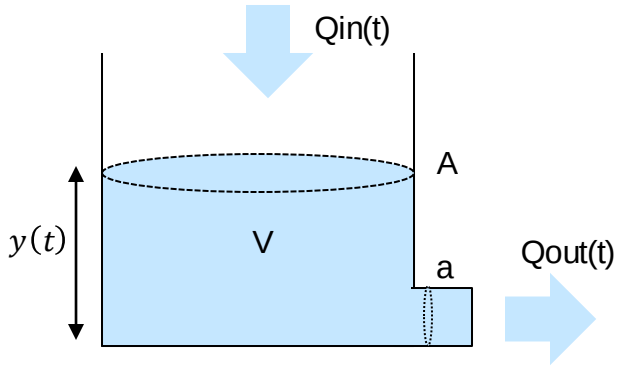


- Rate of change of volume = inflow rate – outflow rate
- $A\frac{d(y(t))}{dt} = Q_{in} - Q_{out} = Q_{in} - f(\mathrm{y(t)}) = Q_{in} - \mathrm{a}\sqrt{2g}\,y(t)$
- $\frac{d(y(t))}{dt} = \frac{G}{A}u(t) - \frac{\mathrm{a}\sqrt{2g}}{A}y(t) = \alpha - \beta y(t)$
- $y(t) = y_0 e^{-\beta t} + \frac{\alpha}{\beta}\left(1 - e^{-\beta t}\right)$

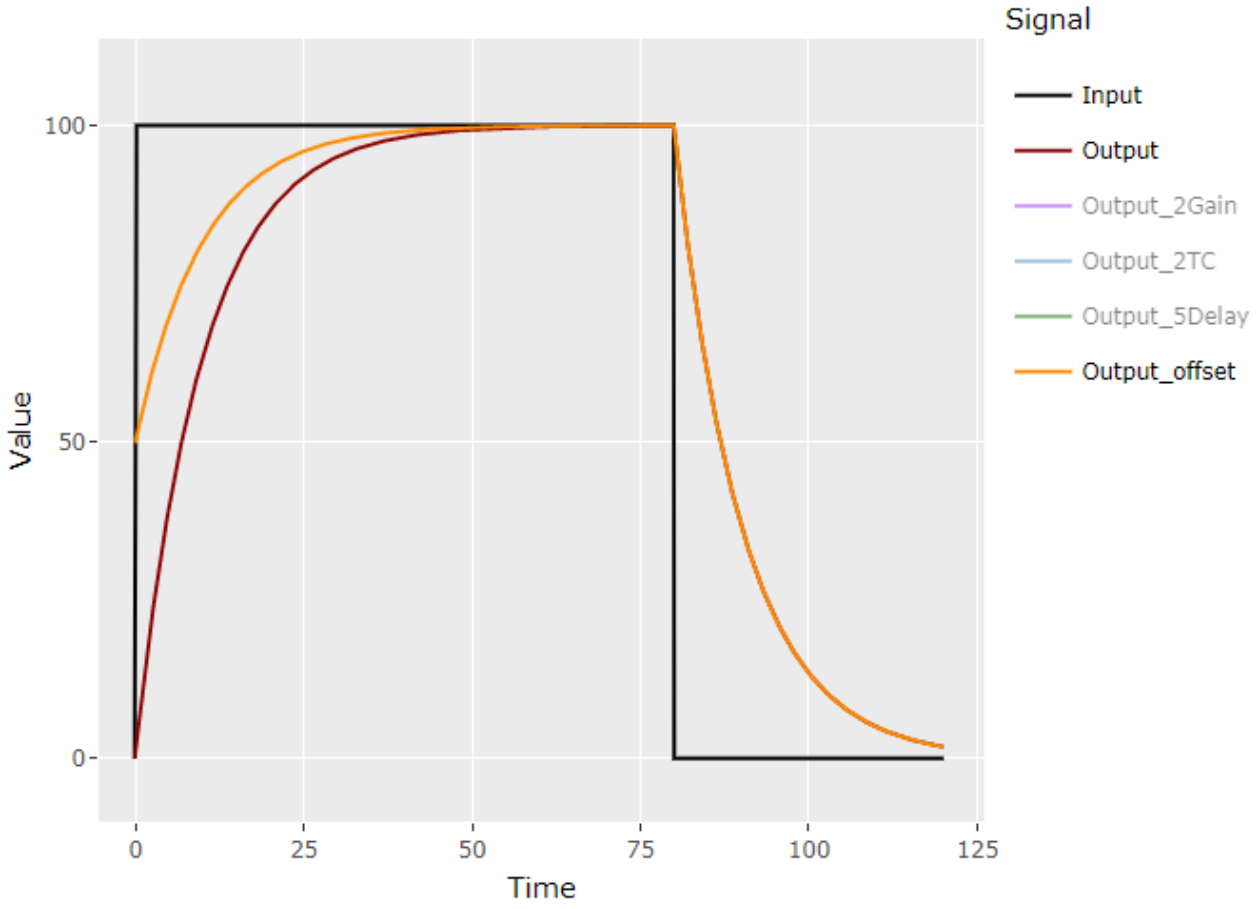


Background – Dynamic Systems

- Leaky bucket (with off switch)

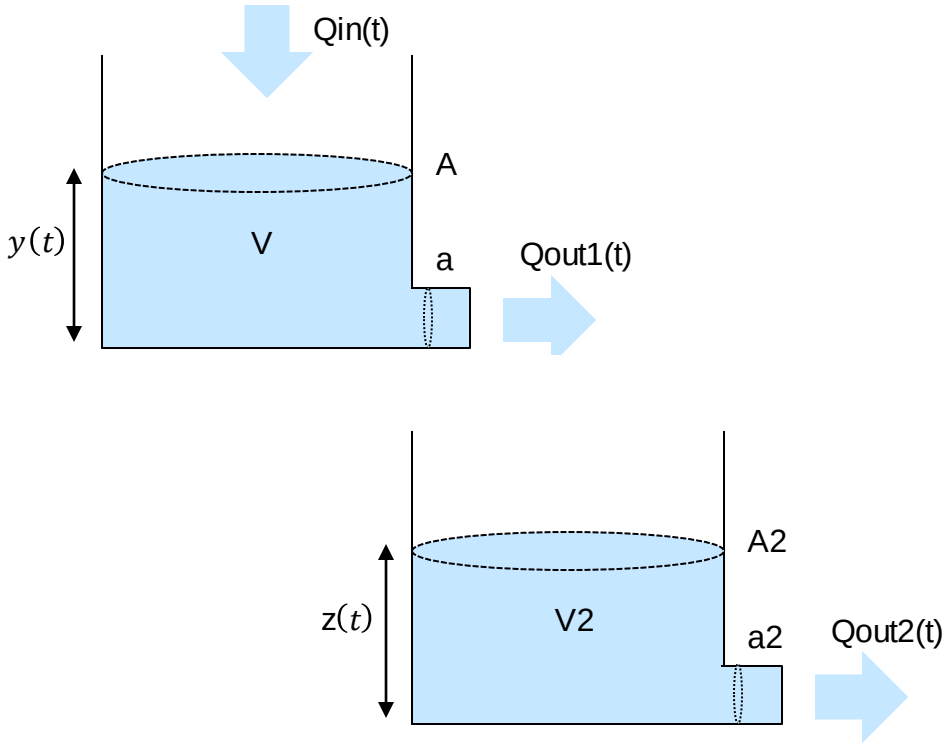


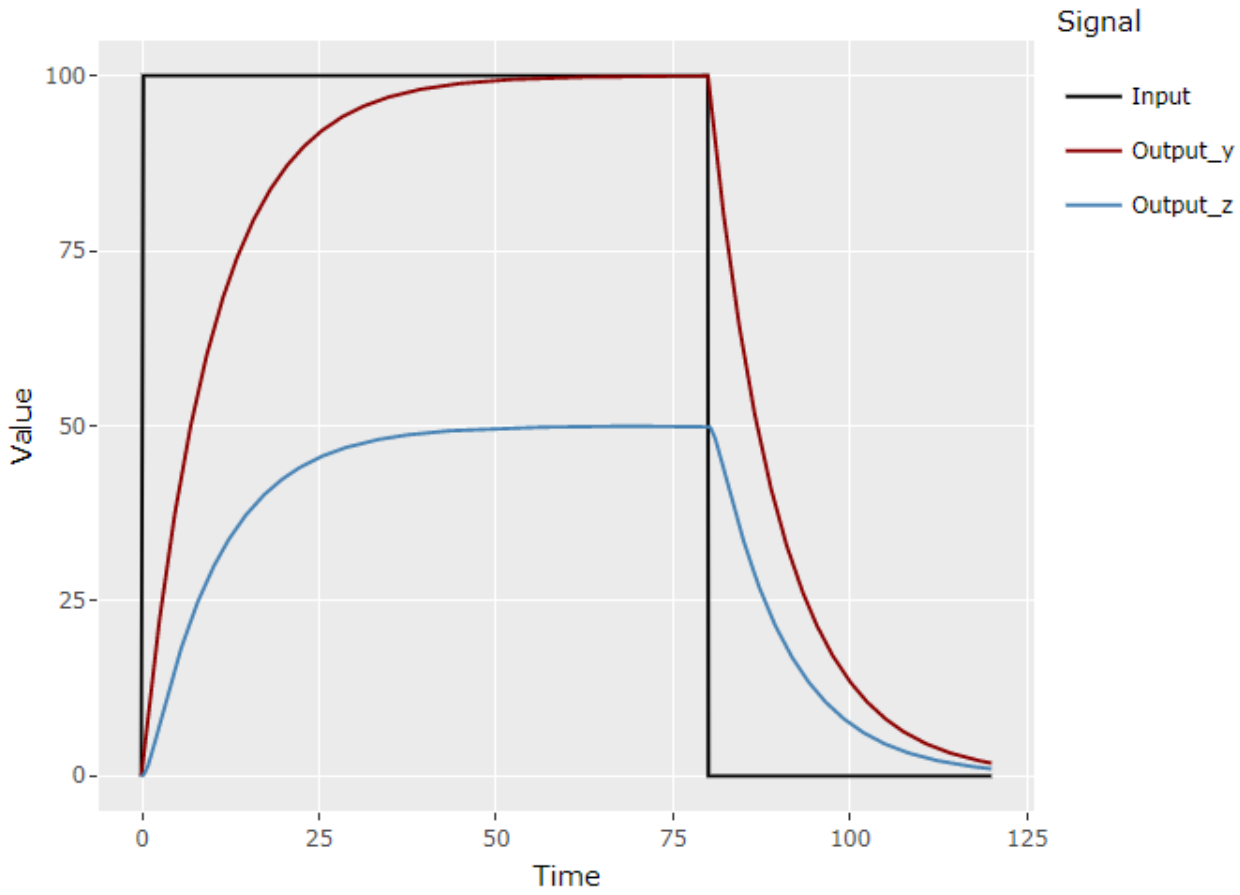
- $A\frac{d(y(t))}{dt} = Q_{in} - Q_{out} = Q_{in} - f(\mathrm{y(t)}) = Q_{in} - \mathrm{a}\sqrt{2g}\,y(t)$
- $\frac{d(y(t))}{dt} = \frac{G}{A}\left(u(t) - u\left(t - t_{off}\right)\right) - \beta y(t)$
- $y(t) = \begin{cases} t < t_{off} : y_0 e^{-\beta t} + \frac{\alpha}{\beta}\left(1 - e^{-\beta t}\right) \\ t \geq t_{off} : \{y_0 e^{-\beta t_{off}} + \frac{\alpha}{\beta}\left(1 - e^{-\beta t_{off}}\right)\} e^{-\beta(t - t_{off})} \end{cases}$



Background – Dynamic Systems

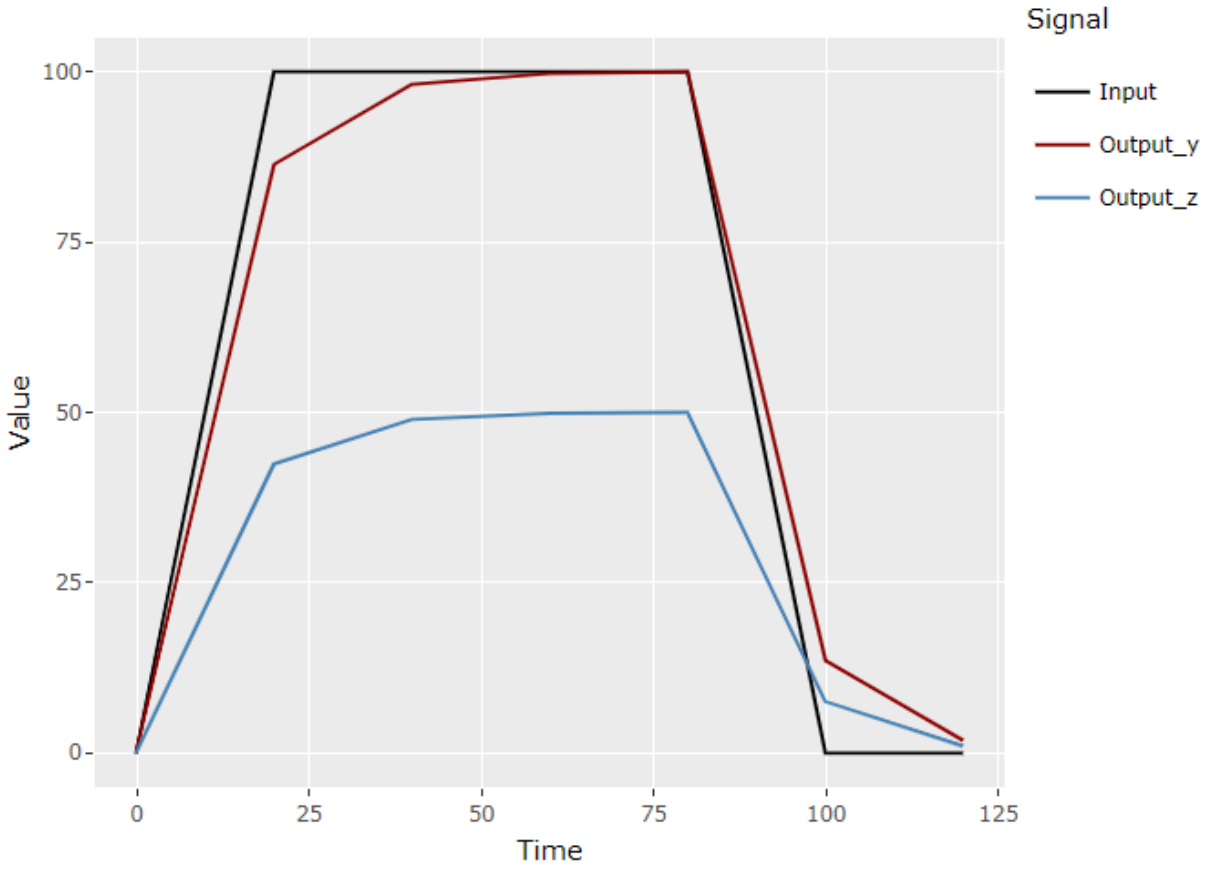
- Leaky buckets



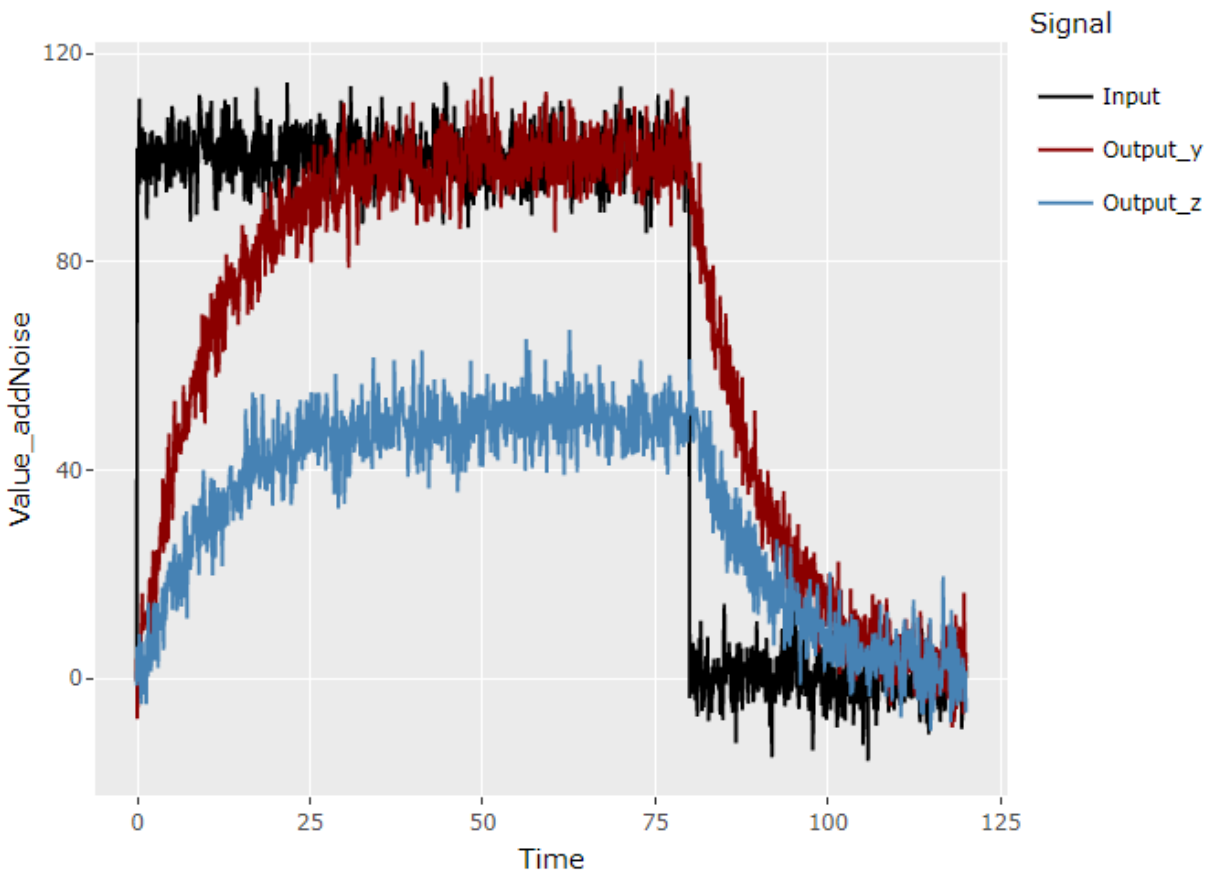


Background – Dynamic Systems - Complications

- Sparsity (Downsampling)

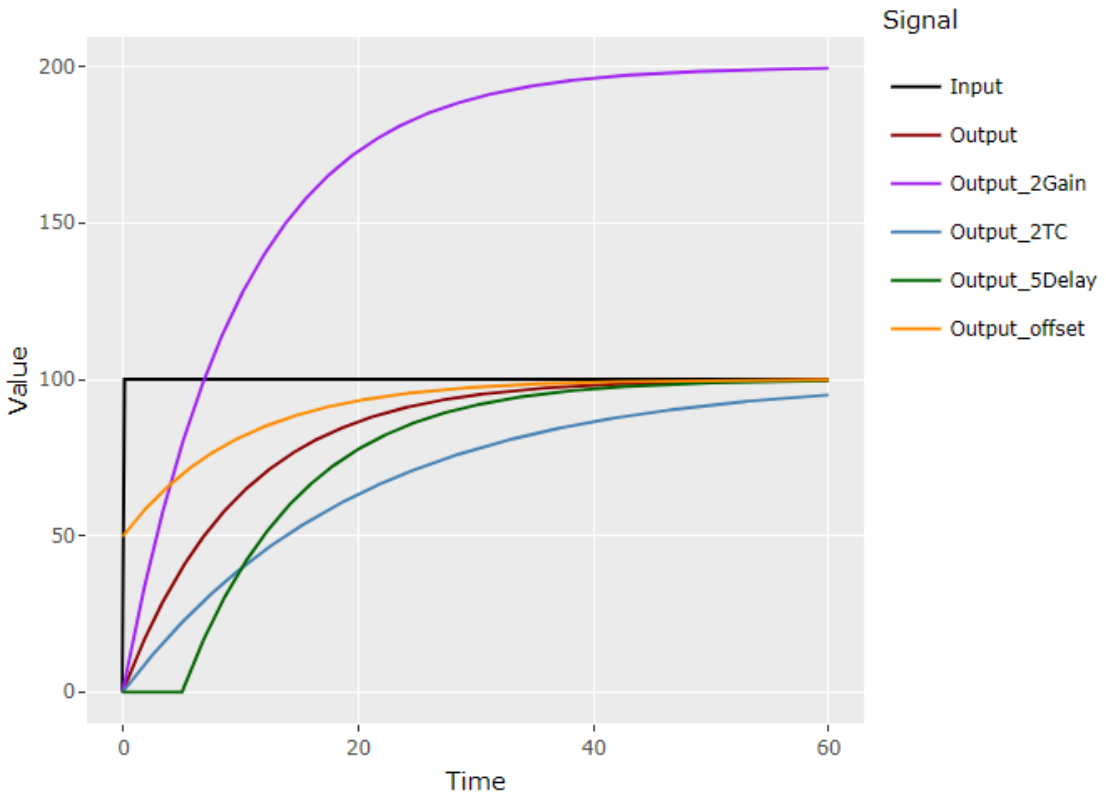


- Noisy (Additive Noise)

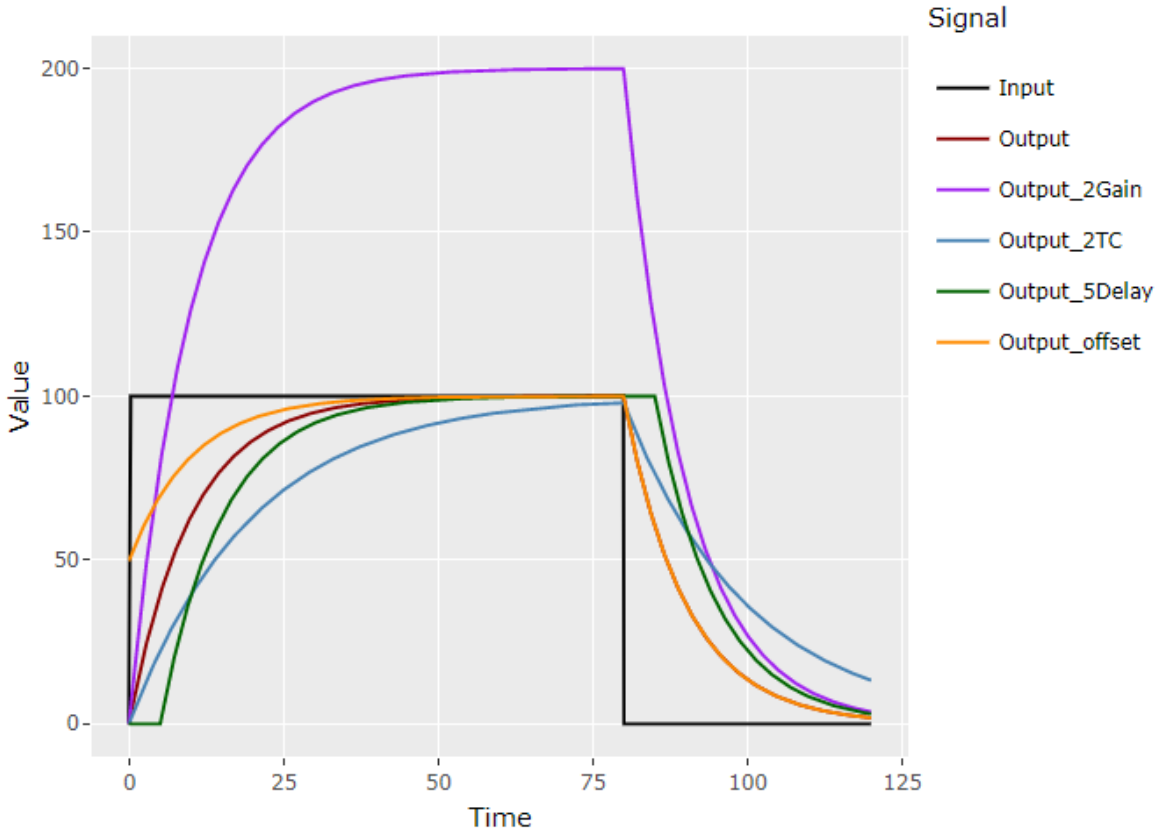


Background – Dynamic Systems: Exploring different Parameters

- Step function responses



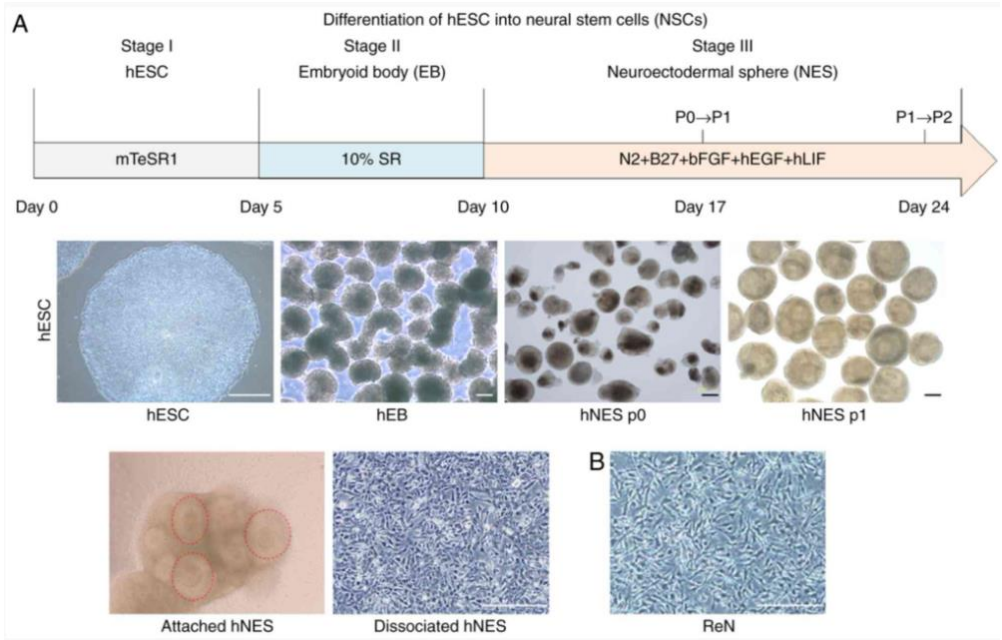
- Square function responses



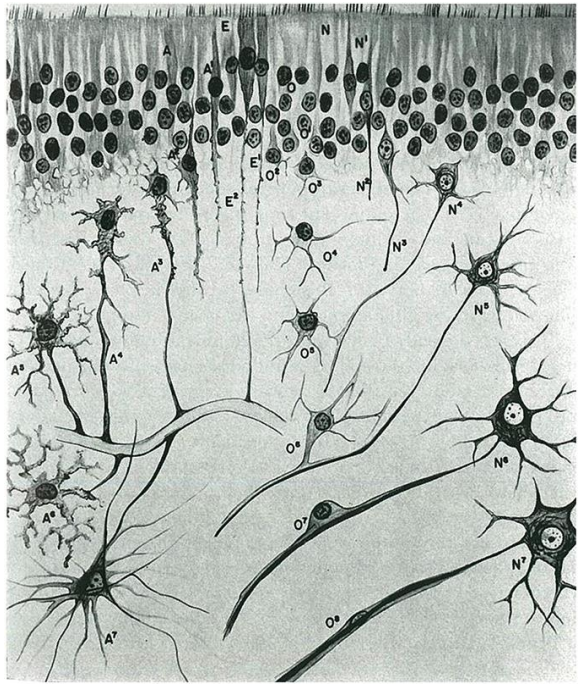
Full Report:
https://github.com/tkamth/Timeresponse/blob/main/UniBasel_DynamicSystem.html

RNA Velocity

- scRNASeq is just a time snapshot but it can capture cells in different states (e.g. proliferation, differentiation)
- Can we infer “cell states” based on transcriptomic profiles?
    - Similar to morphological differences observed from a single histology slide?



https://www.spandidos-publications.com/10.3892/ijmm.2017.3298



Simplified drawing: https://www.frontiersin.org/articles/10.3389/fnana.2016.00013/full

RNA Velocity

LETTER

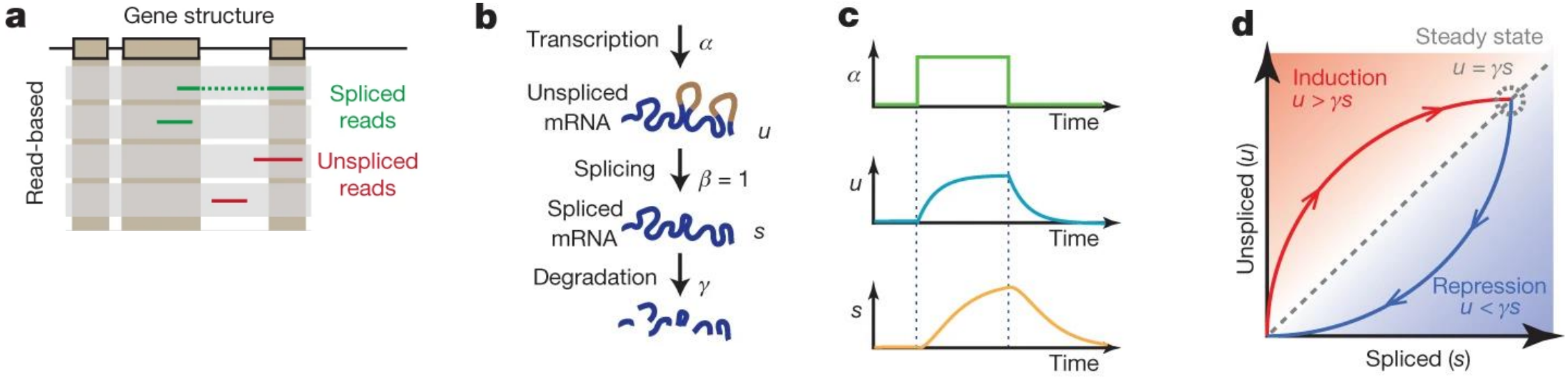
https://doi.org/10.1038/s41586-018-0414-6

RNA velocity of single cells

Gioele La Manno[1,2], Ruslan Soldatov[3], Amit Zeisel[1,2], Emelie Braun[1,2], Hannah Hochgerner[1,2], Viktor Petukhov[3,4], Katja Lidschreiber[5], Maria E. Kastriti[6], Peter Lönnerberg[1,2], Alessandro Furlan[1], Jean Fan[3], Lars E. Borm[1,2], Zehua Liu[3], David van Bruggen[1], Jimin Guo[3], Xiaoling He[7], Roger Barker[7], Erik Sundström[8], Gonçalo Castelo-Branco[1], Patrick Cramer[5,9], Igor Adameyko[6], Sten Linnarsson[1,2*] & Peter V. Kharchenko[3,10*]

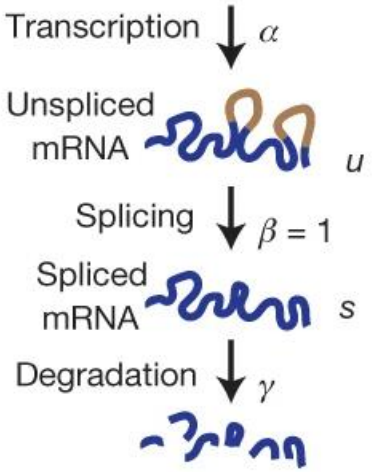
La Manno et al. Nature 2018

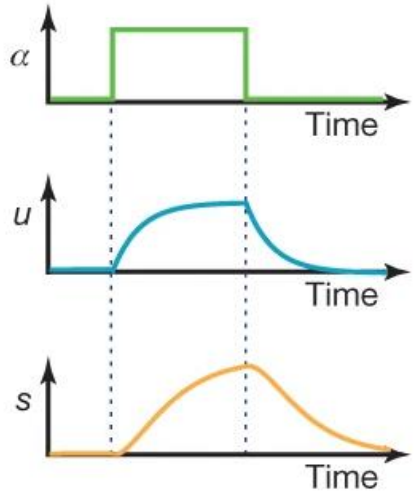
- Velocyto
  - Analysis of expression dynamics in scRNASeq data.
  - Enables estimations of RNA velocities of single cells by distinguishing spliced and unspliced mRNAs
    - Spliced: polyA selection ignores noncoding RNA
    - Unspliced: artifact due to internal priming of intronic sequences



https://doi.org/10.1038/s41586-018-0414-6

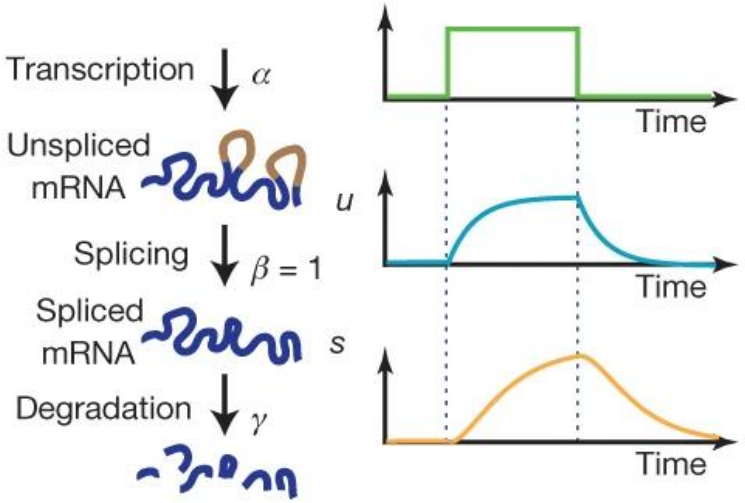
RNA Velocity





- Model transcription as a square wave (On/Off)
- Transcriptional induction for a gene → increase of newly transcribed precursor unspliced mRNAs (until steady state)
- Repression/absence of transcription for a gene → decrease of unspliced mRNAs

RNA Velocity



- System of 1st order Ordinary Differential Equations per gene

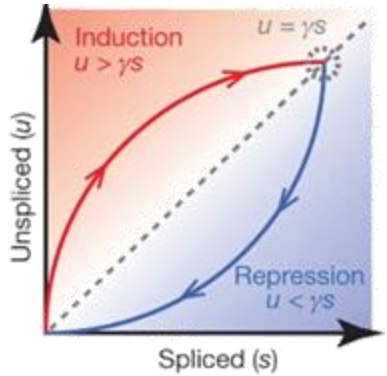
$$\frac{du}{dt} = \alpha(t) - \beta(t)\, u(t)$$

$$\frac{ds}{dt} = \beta(t)\, u(t) - \gamma(t) s(t)$$

- Assumptions
  - time invariant rates
  - common splicing rate across genes (normalizing/measuring rates by units of splicing rate by setting beta=1)

$$\frac{du}{dt} = \alpha - u(t)$$

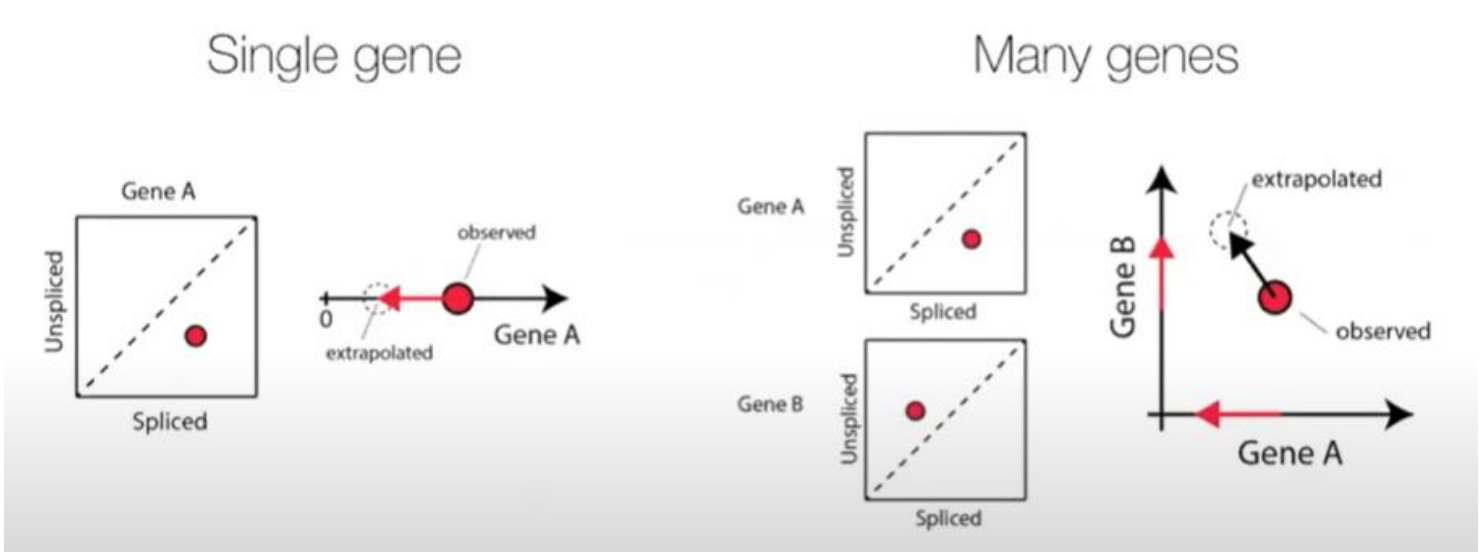
$$\frac{ds}{dt} = u(t) - \gamma s(t)$$

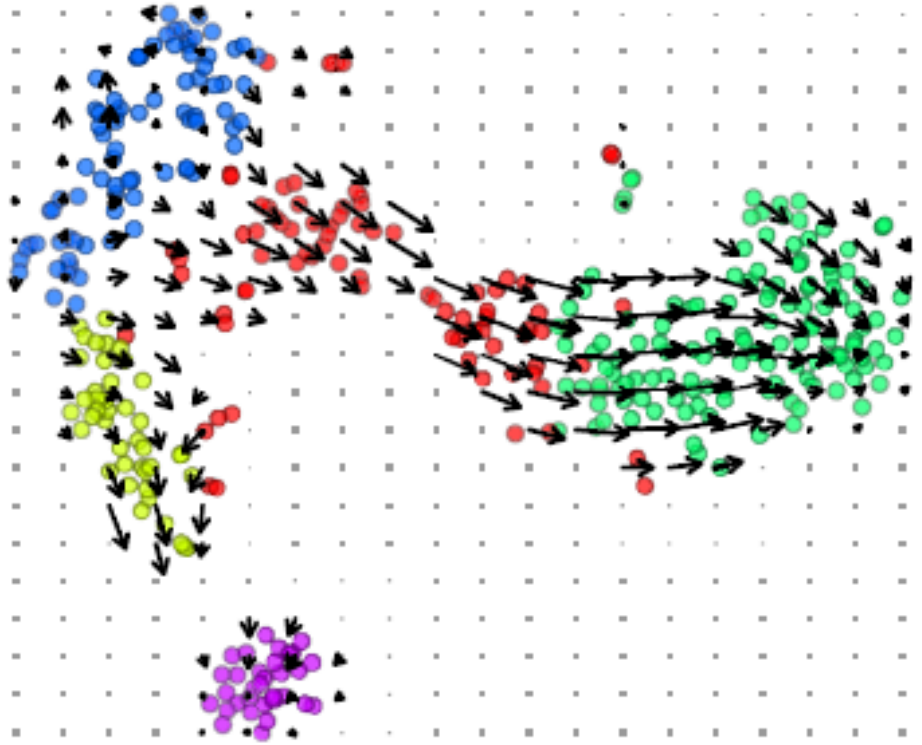


  - Assuming data has cells in steady-state $ds/dt = 0, \longrightarrow \gamma = \frac{u}{s}$
  - Velocity $v = u - \gamma s$ (vertical distance of the observed u from steady state slope)

RNA Velocity

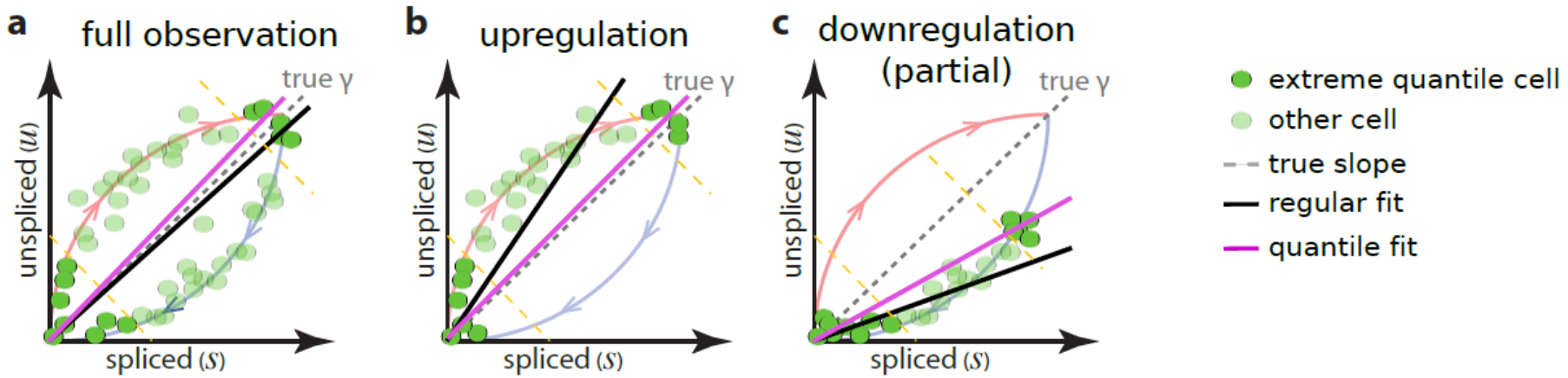
- Velocity (magnitude + direction) from single to multiple genes





RNA Velocity

Achieve more accurate estimation of gene-specific steady-state coefficient γ → regression based on the cells found in the extreme quantiles of expression



Extension

nature biotechnology ARTICLES
https://doi.org/10.1038/s41587-020-0591-3
Check for updates

Roche

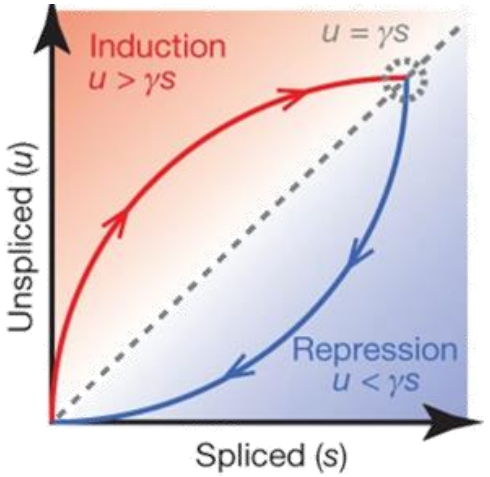
Generalizing RNA velocity to transient cell states through dynamical modeling

Volker Bergen[1,2], Marius Lange[1,2], Stefan Peidli[2], F. Alexander Wolf[1] and Fabian J. Theis[1,2]

Bergen et al. Nature Biotech. 2020

scVelo: Generalizing RNA velocity to transient cell states through dynamical modeling

- Extension to estimate velocity without assuming either presence of steady states or common splicing rate across genes
- Maintains the weaker assumptions of constant gene-specific splicing
- Same transcriptional states are modeled to account for all possible configurations of gene activity
  - 2 steady states
  - 2 dynamic transient states (induction and repression)



https://doi.org/10.1038/s41587-020-0591-3

Extension

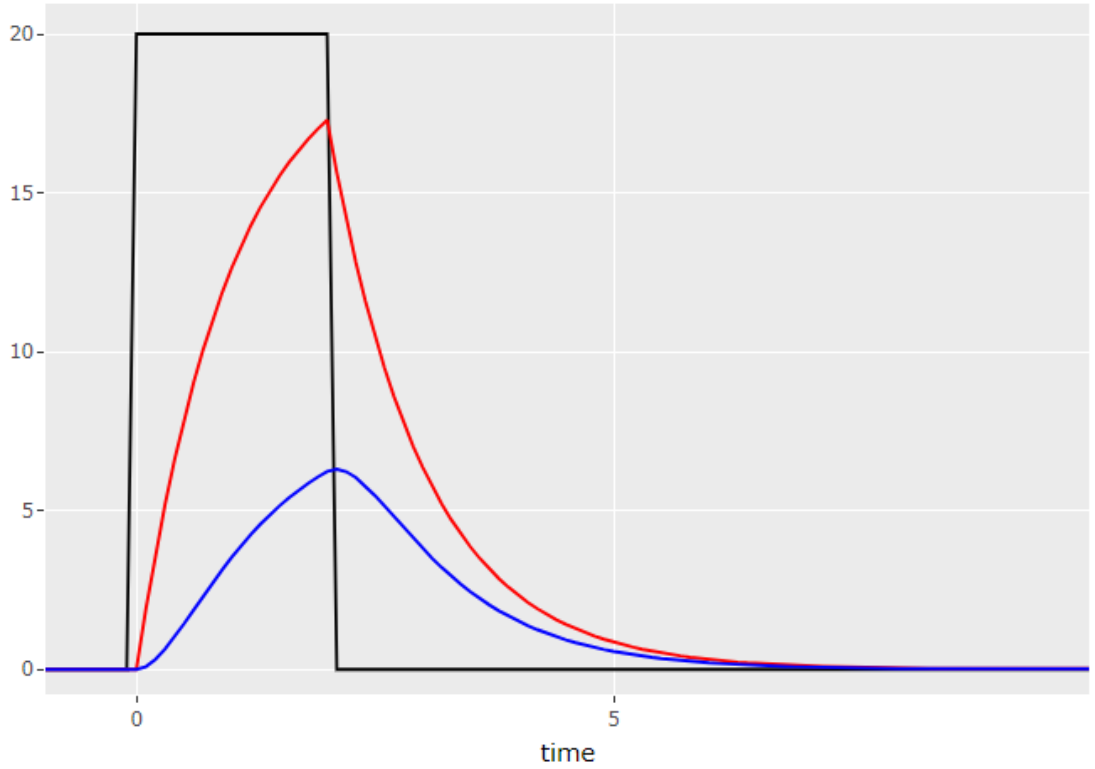
- Dynamic model
  - Define distinct states (induction and regression→ either transcription is on or off)
  - Integrating the differential equations and setting $\tau = t - t_0^{(k)}$ (where k is the state when transcription is either on or off)

$$
\begin{aligned}
u(t) &= u_0 e^{-\beta\tau} + \frac{\alpha^{(k)}}{\beta}\left(1 - e^{-\beta\tau}\right), \\
s(t) &= s_0 e^{-\gamma\tau} + \frac{\alpha^{(k)}}{\gamma}\left(1 - e^{-\gamma\tau}\right) + \frac{\alpha^{(k)} - \beta u_0}{\gamma - \beta}\left(e^{-\gamma\tau} - e^{-\beta\tau}\right)
\end{aligned}
$$

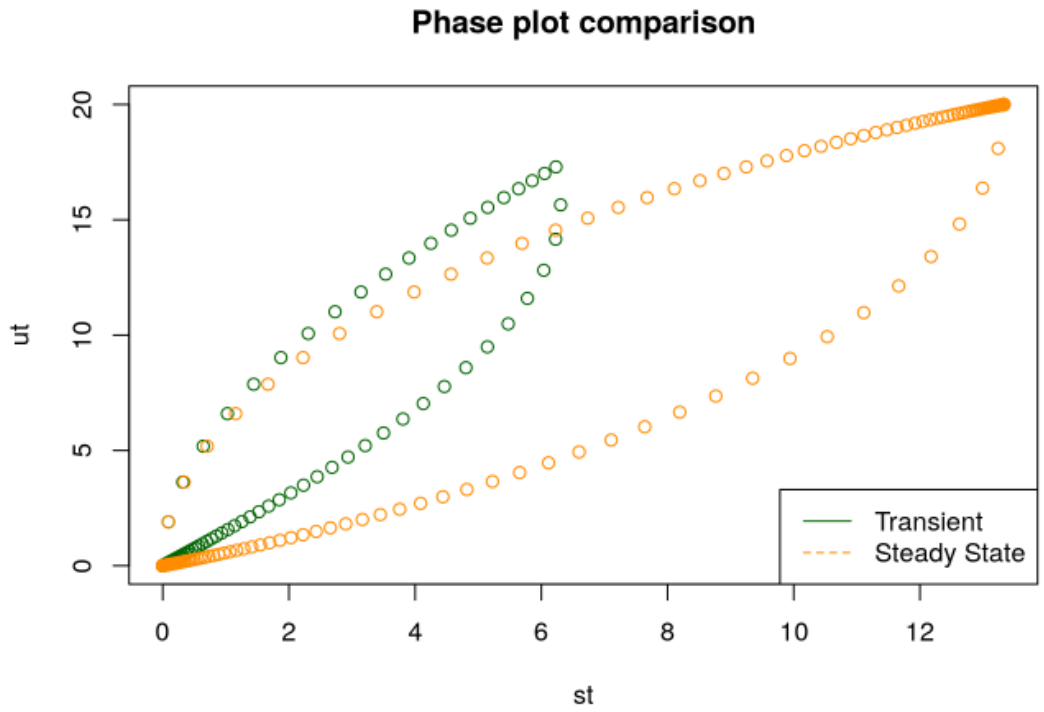
  - Two distinct set of steady states $\left(u_\infty^{(k)}, s_\infty^{(k)}\right) = \left(\frac{\alpha^{(k)}}{\beta}, \frac{\alpha^{(k)}}{\gamma}\right)$

Extension

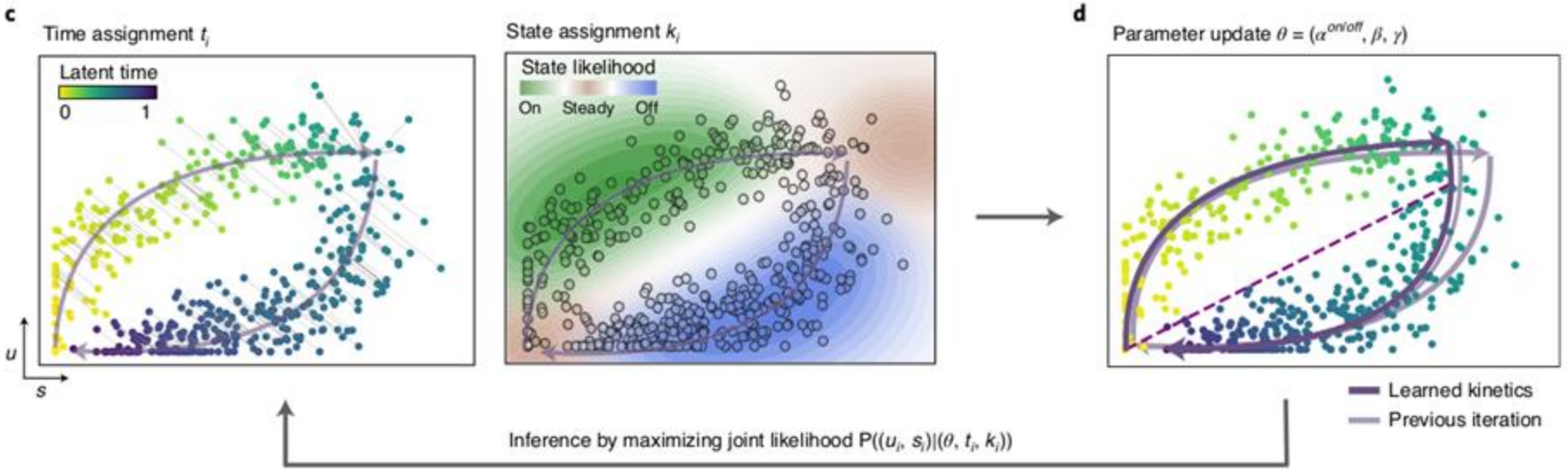
Steady state not reached



Phase plot



Extension-EM



Challenge: both time and rate parameters are unknown

- E step: Assign a latent time to the observed value by minimizing the distance to the phase trajectory model

- M step: reaction rate parameters updated (updated to maximize the log-likelihood using downhill simplex method Nelder–Mead) → new trajectory phase

https://doi.org/10.1038/s41587-020-0591-3

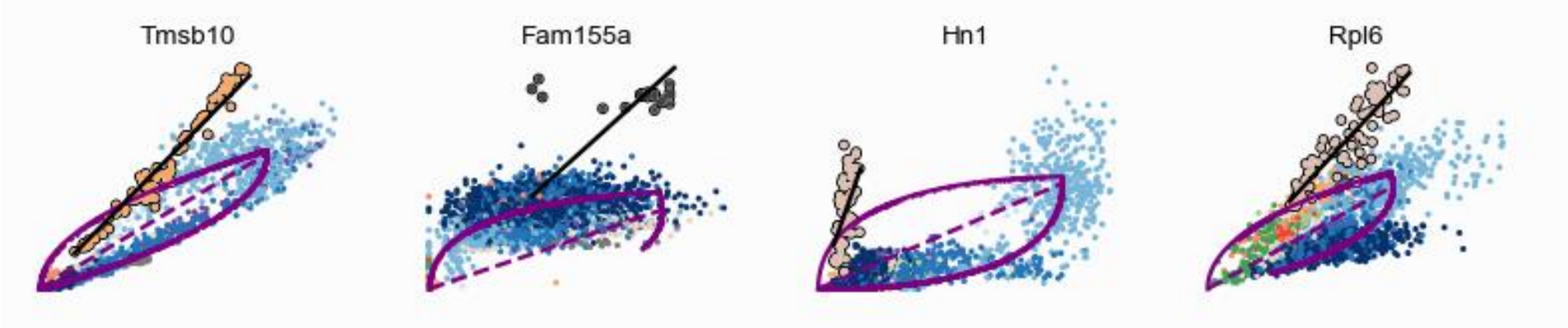
RNA Velocity

- Differential Kinetic Test
  - Distinct cell clusters may exhibit different kinetics regimes

  - Likelihood ratio (asymptotic chi-squared distribution) can be tested for significance.
    - Ratio of a single-kinetic and two-kinetic model (i.e.:supporting multiple clusters),

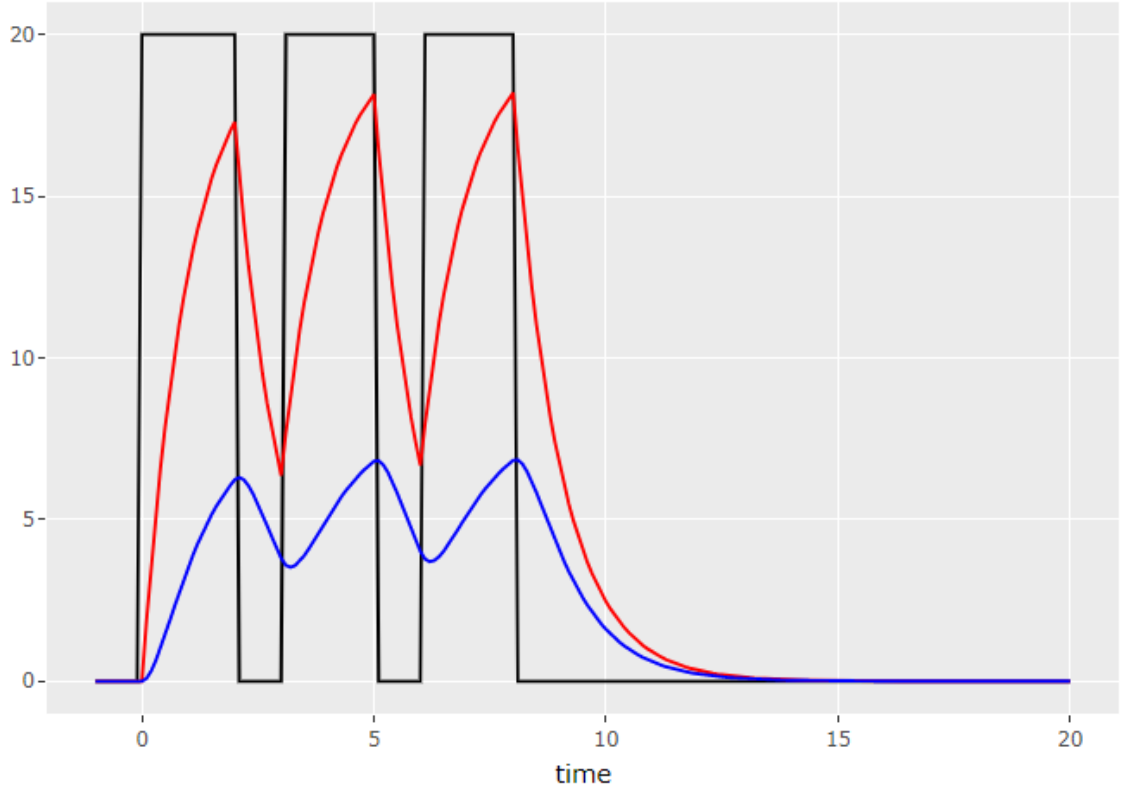
$$LR = -2 \ln \frac{sup_{\theta} \mathcal{L}(\theta)}{sup_{\theta'} \mathcal{L}(\theta')}$$

    - Limitation: For computational reason→ by default an orthogonal regression is used instead of a full phase trajectory to test whether a cluster is well explained by the overall kinetic or not

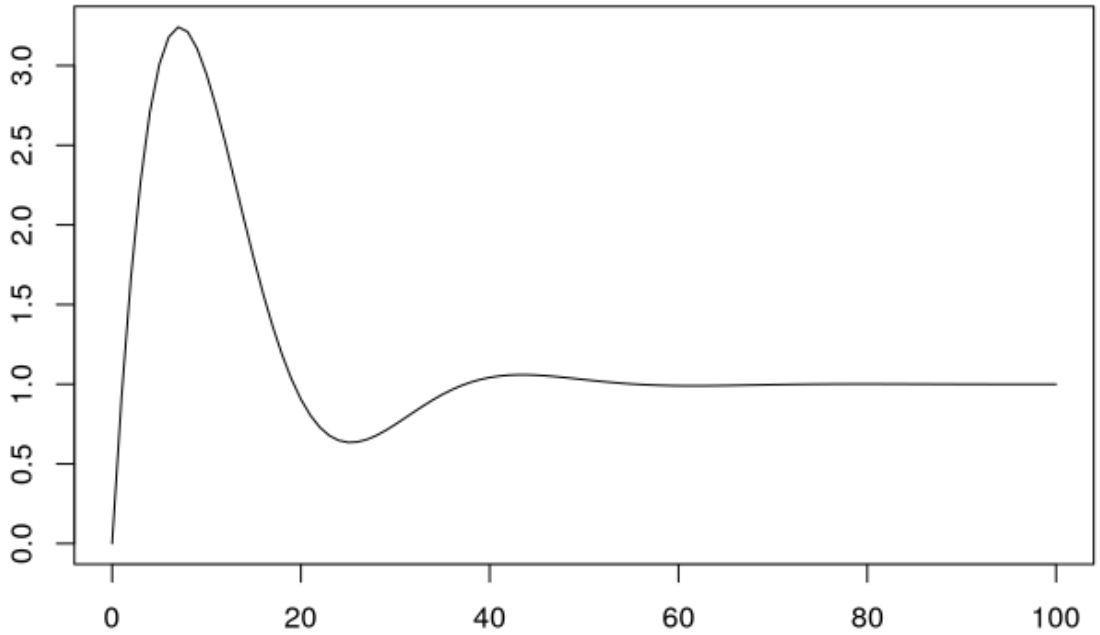


Further Extensions – Complicated inputs and models

Toggling or more frequent switches of transcription input signal



Second order dynamic systems



Theory Summary-RNA Velocity

- Steady state / deterministic model (velocyto)
  - Approximated with a linear regression on the presumed steady states in the lower and upper quantiles.
  - 2 fundamental assumptions:
    - steady-state mRNA levels is captured in the data
    - a common splicing rate across genes
- Extension to a full Dynamic model (using a likelihood-based framework) (scvelo)
  - Solved by likelihood-based expectation-maximization framework

scVelo-Tutorial (Endocrine pancreas data)

- scRNA sequencing of pancreatic cells during embryonic development in mouse

- Elucidate how endocrine progenitors segregate into different endocrine subtypes during development
  - Ideal scenario with developmental data and multiple lineages

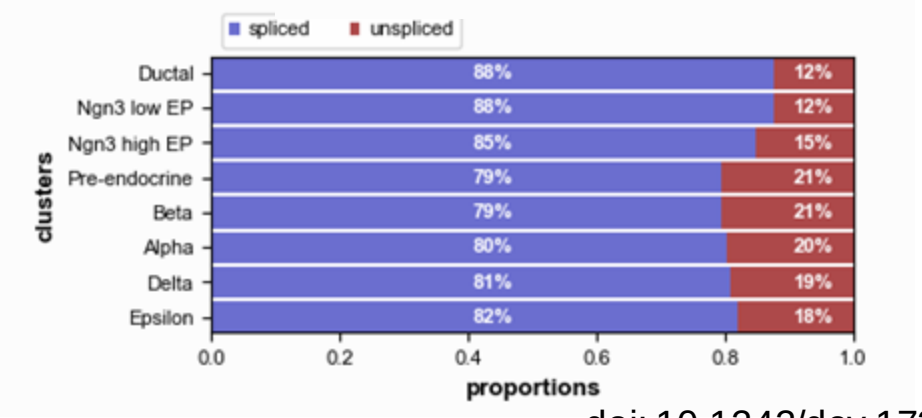
- Typically between 10%-25% of reads are unspliced molecules with intronic sequences

© 2019. Published by The Company of Biologists Ltd | Development (2019) 146, dev173849. doi:10.1242/dev.173849

RESEARCH ARTICLE

Comprehensive single cell mRNA profiling reveals a detailed roadmap for pancreatic endocrinogenesis

Aimée Bastidas-Ponce[1,2,3,4,*], Sophie Tritschler[1,5,6,*], Leander Dony[5,6,7], Katharina Scheibner[1,2,3,4], Marta Tarquis-Medina[1,2,3,4], Ciro Salinno[1,2,3,4], Silvia Schirge[1,2,3], Ingo Burtscher[1,2,3], Anika Böttcher[1,2,3], Fabian J. Theis[5,8,‡], Heiko Lickert[1,2,3,4,‡] and Mostafa Bakhti[1,2,3,‡]

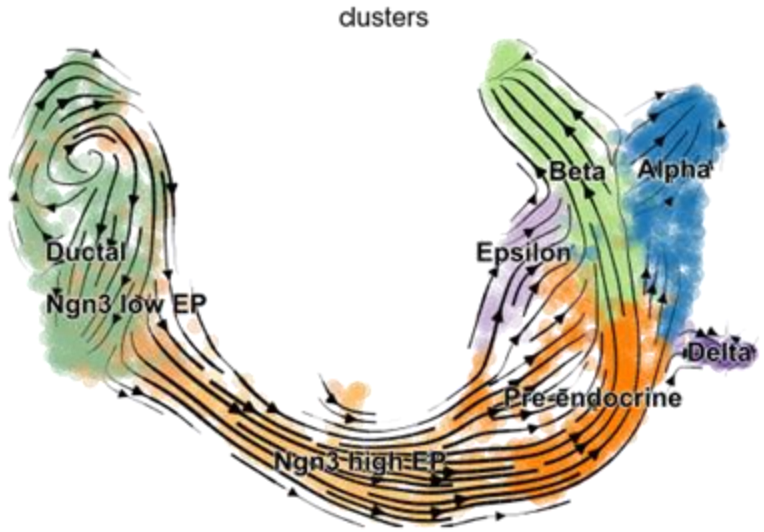


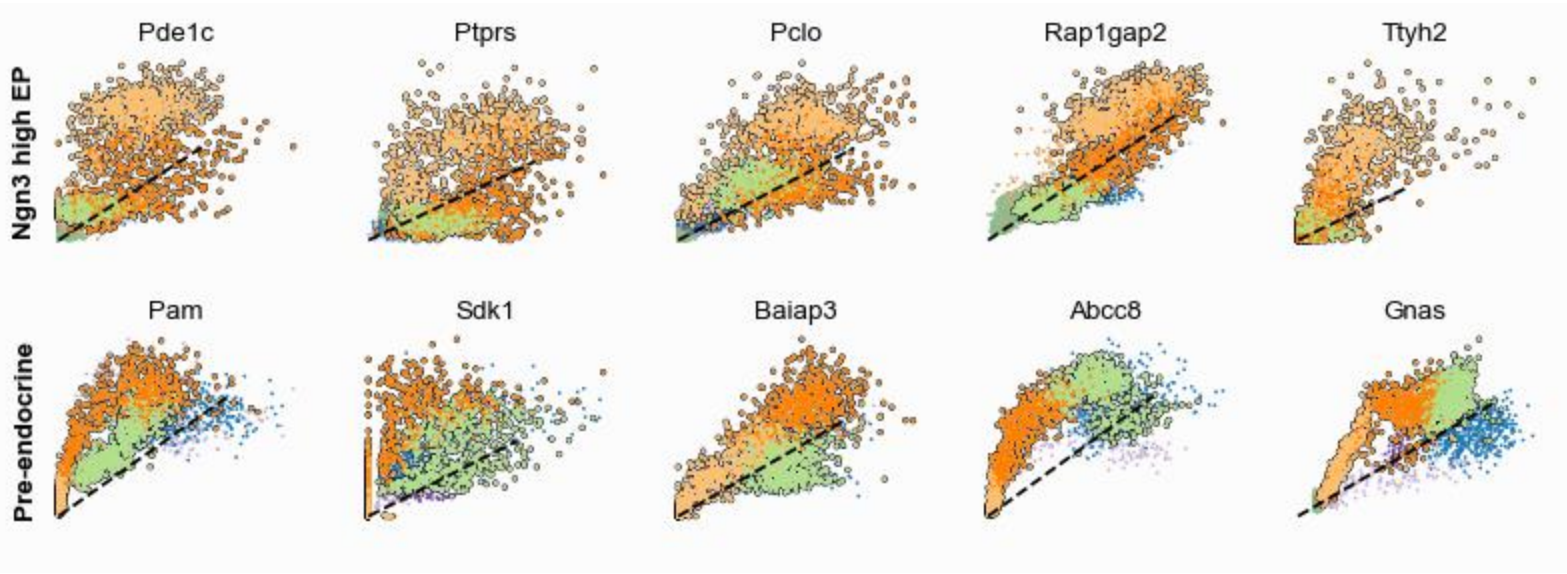
doi: 10.1242/dev.173849

https://scvelo.readthedocs.io/

scVelo-Tutorial (Endocrine pancreas data)

- Cluster specific differential velocity expression in two transient clusters
  - Neurogenin 3 high endocrine progenitor cells (yellow)
  - Pre-endocrine (orange)





Can we target these driver genes to alter the terminal cell states associated with healthy controls?

https://scvelo.readthedocs.io/

CellRank

New Results

Comment on this paper

Previous

CellRank for directed single-cell fate mapping

Posted November 20, 2020.

Marius Lange, Volker Bergen, Michal Klein, Manu Setty, Bernhard Reuter, Mostafa Bakhti, Heiko Lickert, Meshal Ansari, Janine Schniering, Herbert B. Schiller, Dana Pe'er, Fabian J. Theis

doi: https://doi.org/10.1101/2020.10.19.345983

Identifying initial and terminal states

- How likely each cell is to develop from each initial state(s) or towards each terminal state(s)

- Challenge: Interpretation of which genes are supporting the direction of the flow for that cell
  - biased of the umap projection to genes that are more highly abundant

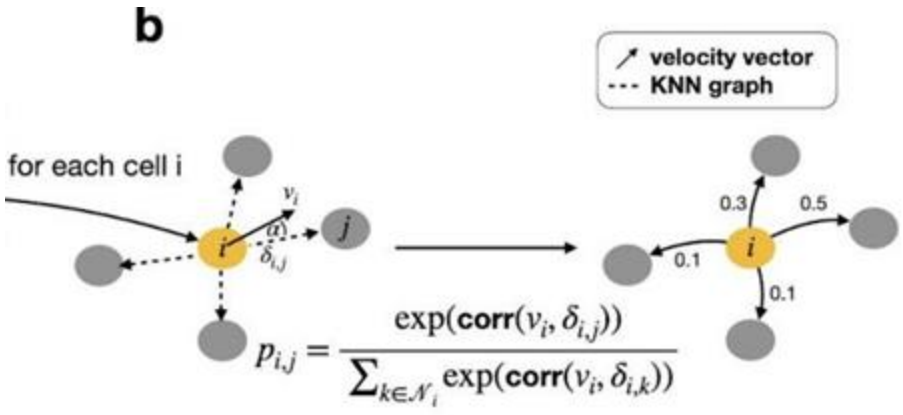
Objective: Construct probabilistic fate maps

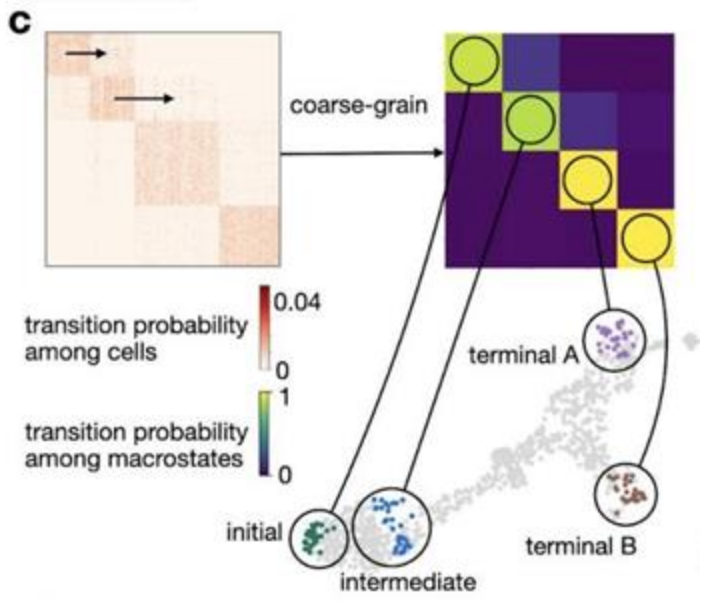
https://doi.org/10.1101/2020.10.19.345983

CellRank

Markov chain transition matrix

- Use each velocity vector to find likely cell transitions that are accordance with that direction
- Transition probabilities are computed using dot product projection between the potential cell-to-cell transitions and the velocity vector and stored in a matrix denoted as velocity graph.



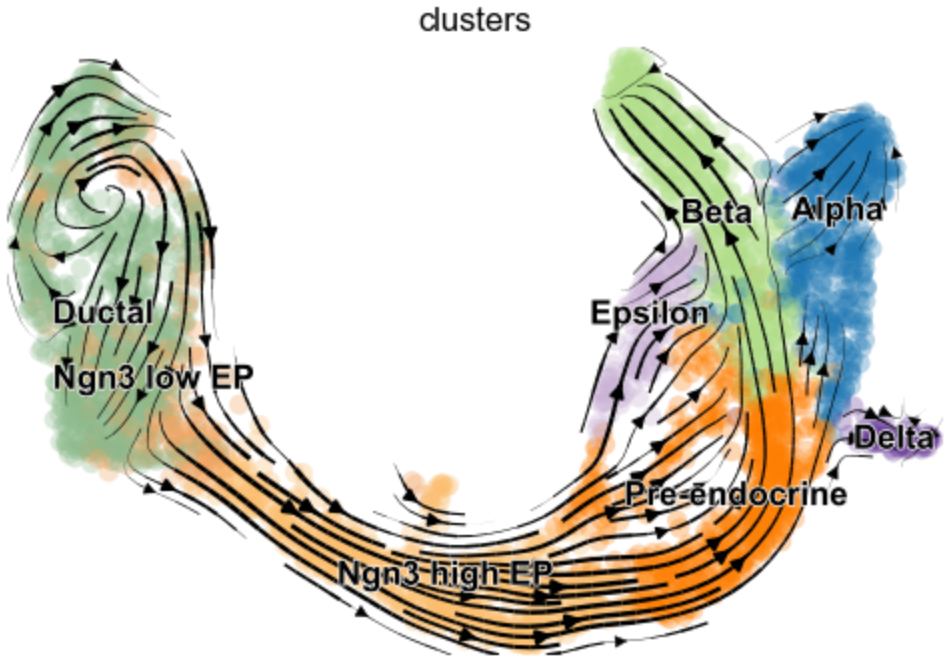


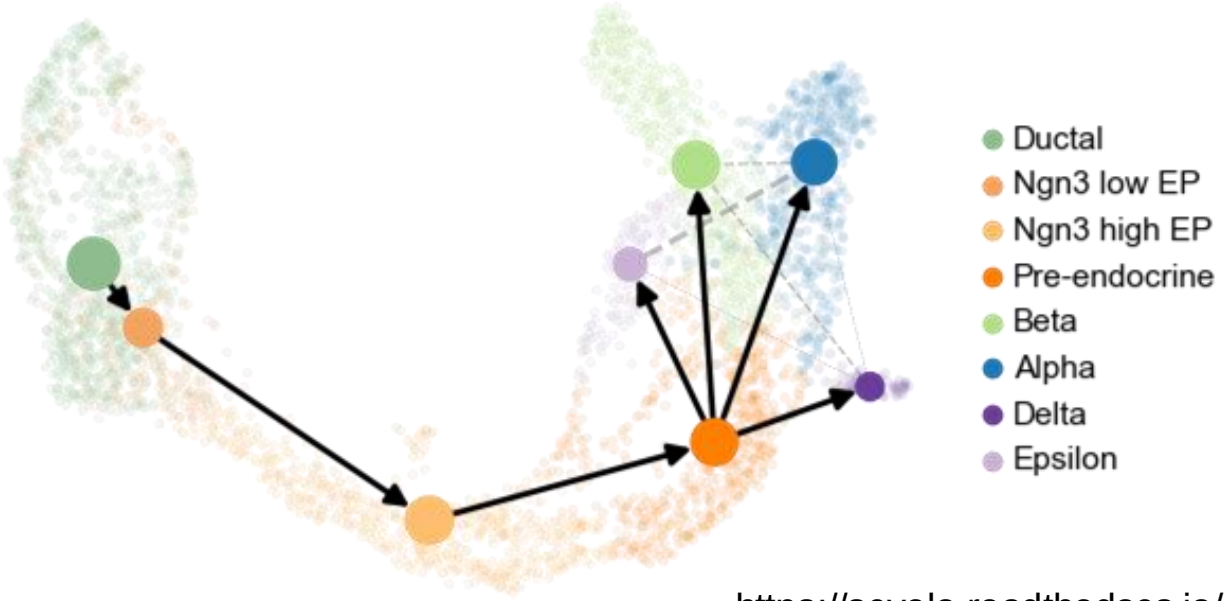
https://doi.org/10.1101/2020.10.19.345983

scVelo-Tutorial (Endocrine pancreas data)

Developmental processes

- Delineates cycling population of ductal cells and endocrine progenitor cells
- Illuminates endocrine cell differentiation and lineage commitment





https://scvelo.readthedocs.io/
https://cellrank.readthedocs.io/en/latest/

Conclusions

Key Messages

- Time series data are fun, powerful and omnipresent
- Data can be sparse and noisy → careful modeling has be factored in
- RNASeq velocity & trajectory analyses provide novel insight into genes important for lineage transitions (*continued effort and assessment by many scRNASeq squad colleagues*)

- Many challenges to model the data correctly while reflecting the true biological mechanism that is often oversimplified
- Novel algorithms and solvers are needed to “integrate” multiscale modelling (computational time)
- Words of wisdom:
  - When you decide to embark on a project → go full speed and deep until you reach “steady state” before the system or/and input signal switch